DOCUMENT RESUME

ED 313 899 FL 018 236

AUTHOR Barnwell, David
TITLE "Naive" Native Speakers and Judgments of Oral Proficiency in Spanish.
PUB DATE 89
NOTE 24p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Interrater Reliability; Interviews; *Language Proficiency; Language Tests; *Native Speakers; *Oral Language; *Rating Scales; *Spanish; *Speech Skills; Testing; Test Validity
IDENTIFIERS *ACTFL ETS Language Proficiency Guidelines

ABSTRACT

A study addressed issues of concern in the use of the American Council on the Teaching of Foreign Languages (ACTFL)/Educational Testing Service (ETS) Language Proficiency Guidelines commonly used in determination of oral language proficiency. Specifically, potential discrepancies between the judgments of trained raters and "naive" native speakers of the target language were investigated. Four recorded oral interviews by an ACTFL-trained interviewer and corroborated by a second trained rater, were played to 14 "naive" (untrained) raters. The experiment deliberately attempted to avoid the process of socialization that plays an important part in the ACTFL training of interviewer-raters. Raters were given a Spanish translation of the generic ACTFL oral proficiency scales and at least one day to study them. The "naive" raters evaluated the recordings based on their own understanding of the scales, without outside interpretation. The results suggest that current knowledge of native speakers is inadequate for prediction of their assessment of non-native speech, even with the use of the scale, raising questions about the generalization of native-speaker reactions as indices of a candidate's proficiency. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED313899

"NAIVE" NATIVE SPEAKERS AND JUDGMENTS OF ORAL PROFICIENCY IN SPANISH

David Barnwell
Department of Spanish & Portuguese
Columbia University
New York 10027

This research was conducted with the aid of funds from the Government of Spain, administered through the Federico De Onís award of Columbia University.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Barnwell, D.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

FL018236

BEST COPY AVAILABLE

"Naive" Native Speakers and Judgments of Oral Proficiency in Spanish

The ACTFL/ETS Proficiency Scales

Since the first training workshops in the use and applications of the ACTFL/ETS oral interview and scales (1982, 1986) a large number of such sessions have been conducted throughout the U.S. Hundreds of persons have been certified to administer the test, and many more have acquired an informal familiarity with the procedure. Many foreign language texts, particularly at elementary level, claim to reflect a proficiency-oriented methodology. The scales have been incorporated into teacher certification programs (Reschke 1985, Hiple and Manley 1987) and have been used by universities as a means of defining entry and exit requirements to their foreign language programs (Arendt, Lange and Wakefield 1986, Freed 1987, Schulz 1988). In many states the proficiency movement has had a significant impact on curricula and testing at the high school level (Cummins 1987).

The ACTFL/ETS oral proficiency interview has thus come to occupy a prominent place in the foreign language teaching and testing methodology of the 1980s. On the theoretical level, the scale has attracted widespread interest among researchers. For some commentators, it has offered to provide an "organizing principle" (Higgs 1984), a unified way of looking at the often divergent procedures and methodologies employed in the foreign language classroom (Omaggio 1986). For Liskin-Gasparro (1984b, p.482) "the value of proficiency tests is that they measure by

definition real-life language ability". In her view they provide "an outside perspective and a check that the instructional goals, methods and outcomes are all synchronized". Elsewhere she expresses a striking confidence in the guidelines, stating that "although problems still remain, they are logistical rather than theoretical" (1984a, p.39). Another advocate of a proficiency orientation (Bragger 1986) shares the conviction that, independent of their validity in testing, the guidelines have produced beneficial change in "curriculum design, teacher behavior, classroom strategies and materials". Buck and Hiple (1984, p.528) assert that "proficiency-based instruction leads to a more efficient, structured curriculum, as well as to increased understanding of and participation in the learning process".

Other commentators, however, have been less enthusiastic. Hummel (1985, p.15), an early critic of the procedure, believes that the guidelines "fail to distinguish between general cognitive skills that are independent of the level of proficiency in the target language and language skills that are related to achievement in the target language". Lantolf and Frawley (1985) criticize the circuitous reasoning embodied in the scales, charging that proficiency levels are defined in terms of themselves. Subsequently, Lantolf and Frawley (1988) have gone even further, and now call for a moratorium on the use of the scale. Others (Bart 1986, Kramsch 1986) have alleged that the proficiency scale lies too much within the "discrete-point" testing tradition. For critics such as these the scale continues to value grammar to the neglect of other components of

communication. Further, they believe that a stress on oral proficiency inevitably leads to a neglect of the many other objectives of foreign language instruction. Bachman and Savignon (1986) argue that the so-called "direct" nature of the test is really an illusion--that it is impossible to divorce testing method from what it sets out to test. Bachman (1988, p.159), though broadly sympathetic to the goals of proficiency testing, charges that ACTFL "has not as yet taken seriously the test developer's responsibility for demonstrating the validity of the interpretation of the ratings and identifying the uses for which they are valid". He warns, as does Magnan (1987) of the difficulty of defending the validity of oral proficiency scores were they to be used in hiring or promotion decisions and consequently subject to challenge in the courts. Generally, there is unease at the lack of a sound empirical base for many of the assertions and assumptions of the proficiency movement, as well as dissatisfaction with the haphazard way in which these notions have been disseminated throughout the profession (Gaudiani 1987).

As can be seen, the early euphoria surrounding the proficiency guidelines has been dispelled by studies which cast doubt on the whole or parts of the procedure. At present there is considerable division within the foreign language profession on the status that should be accorded to the oral proficiency interview. Yet the concept of measurable proficiency is a powerful one, and the ACTFL-inspired proficiency movement shows considerable resilience in the face of its critics. In many

areas it appears that the influence of the guidelines has still to reach its highest point. In this light there remains much scope for research on the proficiency testing procedure.

The Native Speaker

Clearly, the persons with whom one has to interact in the target language are the masses of people who have no training in linguistics or language teaching or testing. Since few people voluntarily study a foreign language in order to talk to their teacher, it is safe to say that most learners aspire to communicating with a wide range of speakers outside the classroom and far from the setting in which they have studied the language. This is probably even true (Morello 1988) of a good number of those American students who take a foreign language as part of a required course of studies.

This assumption, that the native speaker is the target of communicative efforts, is visible in the ACTFL Oral Proficiency Guidelines, in both the 1982 and 1986 versions. Though the 1986 scale less explicitly invokes the native speaker, we still see statements such as "the Advanced level speaker can be understood without difficulty by native interlocutors". Superior level speakers, we are told, may make errors "which do not disturb the native speaker".

Given the use of the "native speaker" as the hypothetical audience for oral interview candidates' efforts to communicate, the role of the ACTFL interviewer/rater is to act as

a kind of surrogate for native speakers, eliciting a sample of language as used to perform certain communicative functions within particular areas, and making the same kinds of judgments, albeit in a more structured and self-conscious fashion, as do native speakers when they interact with foreign learners of their language. The language sample, if properly chosen and evaluated, ought to predict a wide range of performances with native speakers. Yet the relationship between ACTFL ratings and those made by `naive' native speakers has so far escaped serious study, even comment. Think for a moment of the amount and type of training administered to ACTFL raters. However we may view the organization of the ACTFL OPI training workshops, whether or not one is satisfied with their concentration on the `hands-on' and their neglect of theory, it is evident that they represent a process which just about no ordinary native speaker ever undergoes. ACTFL raters, through a process of shared experience and socialization, learn to use their scale in a particular way. In fact, one cannot become an ACTFL certified oral interviewer until and unless one has learned to use the ACTFL scale in this approved way. Native speakers, in contrast, have no training at all in either eliciting speech or rating it. Thus the training of interviewer/raters involves a process of socialization and group identification with an interpretation of the proficiency construct which may be American rather than native speaker in origin (Engber 1987, p.55). In short, the more formal training we give to our apprentice raters the more their experience diverges from that of the native speaker.

ACTFL has insisted on quite a long training period for those wishing to become interviewer/raters. The reason usually advanced for this is that the OPI instrument is a difficult one to administer. Whatever about the validity of that belief in the case of the elicitation component of the interview, there is actually a fair amount of evidence that a long training period is not necessary for the making of reliable judgments on proficiency-type interviews (Frith 1979, Shohamy 1983, Henning 1983, Barnwell 1987). As one recent commentator puts it "assessing communicative effectiveness is not an esoteric skill requiring arduous special training and licensing; it is one of the normal components of linguistic and social adulthood" (Nichols 1988, p.14).

Although there is a dearth of published findings on oral proficiency ratings by 'naive' native speakers, there is a fair amount in the literature on what might be called judgmental strategies followed by native speakers when evaluating the speech of foreigners (Ludwig 1982). This is a very heterogenous body of research, and in no way permits any generalizations about native speaker behavior in any language. A weakness evident in a lot of the work published in the early part of this decade was that the raters of foreigner errors were very often recruited from groups of university or high school students of English, clearly an atypical sample of the general population.

It might be thought that ACTFL would have carried out such research when designing the Oral Proficiency scales, particularly when, as we have seen, the scales contained

several references to native speakers' judgments. Unfortunately, little such research has been forthcoming. When we read references to how the native speaker reacts to speakers of particular levels we are merely dealing with a set of hypotheses rather than observations from the field. Several commentators have stressed the desirability of filling in this gap in our knowledge. Byrnes (1986, p.9) urges research which would involve "obtaining assessments of learner language from native speakers unfamiliar with the rating scale". Clark and Clifford (1987, p.14) call for evidence that "the obtained ratings for given examinees are generally consistent with the judgments of educated native speakers not a priori familiar with this assessment approach". As far as can be seen ACTFL/ETS did not carry out this kind of research when devising the scales, and neither is there evidence that such studies were undertaken on the parent FSI scale, from which the ACTFL/ETS version originates. Indeed, ACTFL has not yet published an authorized translation of the scale to even the most commonly taught languages. Thus we can justly ask how the ACTFL scale can make statements about native speakers when most native speakers of French, German and Spanish could not even read it in their own language. In the absence of detailed work with `naive' native speakers no relationship has been established between how ACTFL raters think and how ordinary natives think. What if we were to find patterns of disagreement between native speakers and ACTFL raters ? Who do we believe if we find that native speakers are consistently more generous, or more strict, than ACTFL raters ?

An Empirical Study

In an effort to begin to find some answers for questions such as these, a study was recently undertaken in Barcelona, Spain. Four tape-recorded oral interviews, carried out by an ACTFL-trained interviewer, whose ratings had been corroborated by a second trained rater, were played to a total of fourteen "naive" raters. In order to see whether the scales expressed any true psychological reality, in themselves and on their own merits, rather than as expounded and explicated by other parties, it was decided to keep training to a minimum. Thus the experiment deliberately set out to avoid the process of socialization and group interaction which plays an important part in ACTFL training of interviewer/raters.

Each rater was issued a translation to Spanish of the generic ACTFL Oral Proficiency scales. In the absence of an officially endorsed Spanish version, the translation had been undertaken by the researcher in collaboration with a native speaker of Spanish. In addition, the raters received a brief description of the aims and strategies of the oral interview, a translation to Spanish of the "situations" which the (ACTFL-trained) interviewer had required her candidates to perform, and an explanation of a small number of culture-specific terms (color guard, fraternity etc.) which came up during the interviews. The raters were then given at least one day to study the scales at home.

The raters subsequently met with the researcher in order to carry out the evaluation. Two sample tapes were played as a pre-experimental exposure to the interview setting and procedure. The purpose of this was not to give the raters benchmarks for particular levels, but merely to accustom them to such things as the interview format and duration, the personality and accent of the interviewer, the kind of conversational topics likely to be raised etc. Having listened to the two trial tapes, the raters embarked on the process of rating each of four oral interviews. All tapes were heard in the same sequence, and no discussion of the tapes was permitted until the entire sequence had been played and written ratings made. Hence the ratings were made independently, the raters basing their judgments on their own understanding of the ACTFL scale, not on how the scale might have been interpreted for them by seasoned users or trainers.

Analysis of Ratings

There are two ways in which raters' judgments can be compared. Firstly, the Spanish judges' ratings can be compared among themselves, to see to what extent they agreed in their assessment of candidates' proficiency. Secondly, they can be set against the ratings made by the ACTFL-trained interviewer, the latter's ratings having been corroborated by another person who had undergone ACTFL training. In this way a comparison can be drawn between the interpretation of the scale reached by "naive" natives and that made by persons trained to use the scale.

Generally, the raters follow similar patterns in tracing the comparative merits of the four candidates. In other words, they agreed that candidate III was the best, that candidate I was second best, that candidate II was next, and that candidate IV was the weakest. This is a not altogether trivial finding, since it tends to demonstrate the validity of the process--that the raters were measuring the same thing. Further evidence for this is rendered by the fact that there was far from a random scatter of ratings in each case--there was clear evidence of patterns in rating.

However, there were quite significant discrepancies as to how the Spanish raters translated their perceptions of the ability of the interviewees into the terms of the ACTFL scale. The first interview to be rated, candidate I, received ratings at five different points, on two different grand levels, Intermediate and Advanced. Ratings for candidates II and III span four points, across two grand levels. Only for candidate IV was there substantial agreement, and this is probably because the raters could not go any lower than Novice Low. As has been seen, though raters agreed on which were the good candidates and which were the weak, they differed substantially as to how to translate their perceptions into points on the ACTFL scale. The same definition was applied to two candidates of apparently very different levels of ability; two candidates of roughly equal ability could be placed very far apart on the scales. For instance, both candidates I and II received some Intermediate-Low ratings, even though candidate I was, in the majority view, far superior to candidate II.

Yet, while pointing out discrepancies such as these, it is worth stressing that about half the Spanish ratings concur in all four cases. Further, in about one-third of all cases the untrained "naive" natives were in exact agreement with the trained rater. Such agreement rates are by no means catastrophically low; whether or not they are adequate is a matter for discussion within the context of the use to which results on the test are to be put. The fact that there is a rather low degree of inter-rater reliability for untrained raters is by no means surprising. Further training, consultation and feedback could be expected to radically improve reliability. A worthwhile subject for a future investigation by ACTFL would be to administer their training experience to a group of "naive" natives. These would then offer a _via media_ between the academic testers in the U.S., on the one hand, and the totally untrained native speaking population on the other. Of course, the more such training is administered to raters the less "naive" they become, raising an interesting psychometric heresy, that of the possibility of a certain tension between validity and reliability in foreign language testing. Leaving aside this speculation, it appears from the present study that the operational descriptions for each ACTFL/ETS level are not in themselves self-sufficient or self-explanatory--they can mean different things to different people. The scale's limpid logic is not immediately visible to the untrained eye.

The statistical treatment of non-parametric ratings or verbal labels presents a special problem in seeking to analyze data. Studies with the FSI scale have traditionally assigned numerical values to ratings, thus permitting correlation coefficients to be worked out. However, such a tactic begs the question of what numerical value to allot to each verbal level. Since no work has been published on how to assign numerical equivalences to points on the ACTFL scale, in order to compare the group ratings with those of the interviewer/rater it might be safest to select the modal rating for each candidate. Thus, for candidate I, the modal rating is Int-High, for candidate II it is Novice-High, for candidate III it is Advanced, and for candidate IV it is Novice-Low. Comparing these with the interviewer/rater's ratings of A+, N-h, S and N-m, it is clear that the naive natives are exhibiting a tendency to be more severe in their judgments than was the interviewer. Looking at the data another way, of 56 paired judgments formed by a comparison between the interviewer's rating and that of each individual "naive" native, 34 (61%) showed the naive native to be more severe, 18 (32%) showed agreement between interviewer and naive native, and in only 4 cases (7%) did the naive native prove more lenient.

This finding is somewhat surprising, since it runs counter to the belief, a belief which has some empirical support (Galloway 1980), that "naive" native speakers are more lenient in their judgments than are professional teachers or testers. Since both the present study and studies such as Galloway's are small

in scope, no attempt can be made at this time to settle the question. We are clearly at a very early stage in the study of native speaker reactions to non-native speech. One hypothesis at least might explain the Spanish raters' reluctance to concede the rating of Superior. It may be that the invocation of the native speaker as the paragon towards which foreigners should strive is more complicated than might first appear. There are scattered suggestions in the literature that native speakers in some cases react negatively to foreigners whose language proficiency is truly of a high order. Loveday (1982) found that Japanese native speakers view adversely those foreigners (Caucasians) who speak Japanese well. Those speaking halting Japanese, on the other hand, are praised and flattered. Loveday observes that the statement is often heard that a Western face and colloquial Japanese do not go together. Gannon, (1980) gives some corroborating anecdotal impressions from his experience in Canada. He asks whether language teachers and testers really know how society reacts to the foreign learner's use of certain types of idiomatic language, for example, or of certain very informal registers of speech. A somewhat parallel finding is discussed by Valdman (1987, p.140), who reviews a study that showed that a group of native speakers of English disapproved of foreigner speech which, though exhibiting a high level of proficiency, bore the influence of regional English dialect. Though these studies do not amount to a convincing body of evidence, they raise the suspicion that rater behavior when faced with the higher levels of proficiency can be less

predictable than is assumed by an easy acceptance of the native speaker as the ideal to emulate. As Nichols (1988, p.15) points out, standardized proficiency scaling presupposes "uniform, incremental, and monotonic increases in the ability to speak a language". Such a model may not correspond to the sociolinguistic reality of communication.

The raters who volunteered to take part in this experiment received a small honorarium for their services. The group was not a random sample of the population, since it was biased towards those with a university background, especially towards psychology and philology students. A proper set of studies of naive native speakers would have to face the problem of how to include a wider cross-section of raters, including those who would not ordinarily volunteer to take part in psycholinguistic experiments.

No formal attempt was made to elicit raters' reaction to the oral proficiency interview and scale. However, from some informal comments it appeared that raters generally viewed the interview positively, considering it to a be a fair sample of candidates' ability, and feeling that it permitted them, the raters, to have an interesting insight into the culture of a far-off land. Many of the raters appeared to enjoy listening to the candidates' efforts to express themselves in Spanish. Impressionistically, raters tended to use the word "fluency" when speaking about candidates. They were somewhat vague in defining what they meant by the term, but as far as they were

concerned, fluency was an important criterion when they judged candidates. On the negative side, several of the raters offered the opinion that the interviews were unnecessarily long. They believed they could form an accurate evaluation well before the interview had come to a close. In addition, many expressed some puzzlement with the use of "situations", offering the view that they would have had no idea of what was going on if these situations had not been explained to them beforehand. And it was noticeable that raters were genuinely lost if a weaker candidate used English words, such as "Placement Office" or "roommate". It bears repeating that the ACTFL scale was conceived and is now used in a particular setting, the U.S university environment, where the majority of language learners are Anglophone monolinguals of a certain age-group. Because interviewer/raters are themselves part of this world, and are quite familiar with its vocabulary, culture, and general assumptions, they may attach insufficient gravity to those occasions when candidates are unable to communicate or express these concepts.

Conclusion

This study confirms the belief that our knowledge of native speakers is at this stage quite inadequate to allow us to predict how they will assess non-native speech, even when they are using a rather elaborate scale such as ACTFL's. It raises the question of whether any proper generalizations can be made about native speaker reactions, and, accordingly, whether a scale can validly cite hypothetical native speaker judgments as

indices of a candidate's proficiency. Yet there remains an epistemological need to involve native speakers in the elaboration of scales which use them as criteria, and their input can only serve to clarify our notions of proficiency and strengthen our proficiency tests. But the more we go out into the real world, the more we involve native speakers, with all their differing attitudes, personalities, prejudices and idiosyncrasies, the more problematic will be the use of any blanket native speaker norm. As of now, the ACTFL/ETS scale's invocation of the concept of native speaker is of unproven validity.

Table I: List of ratings (14 raters, 4 candidates)

Candidate:	I	II	III	IV
Rater:				
A:	I-h	N-h	A	N-l
B:	A+	N-h	S	N-m
C:	I-h	N-h	A	N-m
D:	I-h	N-h	A	N-l
E:	I-l	N-m	A	N-l
F:	I-h	N-m	A	N-l
G:	I-m	N-h	A+	N-m
H:	I-h	I-l	A	N-m
K:	I-h	I-l	A+	N-m
L:	A+	I-m	S	N-m
M:	A+	I-m	S	N-l
N:	A	N-m	A	N-l
O:	I-h	N-m	A+	N-l
P:	I-l	N-h	I-h	N-l
Trained Interviewer/ Rater	A+	N-h	S	N-m

'Naive' Raters

Average age: 26 Age range: 19-44

Males: 6 Females: 8

University students: 7 University graduates: 5

No university education: 2

Abbreviations

N; Novice
I: Intermediate
A: Advanced
S: Superior

h:high
m:mid
l:low

Table II: Analysis of 'Naive' Ratings

(number of times candidates received a particular rating)

Candidate:	N-l	N-m	N-h	I-l	I-m	I-h	A	A+	S
I:				2	1	7	1	3	
II:		4	6	2	2				
III:						1	7	3	3
IV:	8	6							

References

American Council on the Teaching of Foreign Languages Proficiency Guidelines 1982, 1986. Hastings-on-Hudson, New York.

Arendt, J, D Lange and R Wakefield. 1986. Strengthening the language requirement at the University of Minnesota: an initial report. Foreign Language Annals 19, 2, 149-56.

Bachman, Lyle. 1988. Problems in examining the validity of the ACTFL oral proficiency interview. Studies in Second Language Acquisition, 10, 149-163.

Bachman, Lyle, and Sandra Savignon. 1986. The evaluation of communicative language proficiency; a critique of the ACTFL oral interview. Modern Language Journal, 70, 380-90.

Barnwell, David. 1987. Who is to judge how well others speak ?: an experiment with the ACTFL/ETS scale. In Proceedings of the Eastern States Conference on Linguistics, Ohio SU., 37-45.

Bart, Benjamin F. 1986. A note on proficiency and communication in language teaching. ADFL Bulletin, 17,3: 61-62.

Bragger, Jeanette. 1986. Teaching for proficiency: are we ready ? ADFL Bulletin, 18, 1, 11-14.

Buck, Kathryn, and David Hiple. 1984. The rationale for defining and measuring foreign language proficiency programs for business. Foreign Language Annals, 17, 5, 525-8.

Byrnes, Heidi. 1986. Proficiency: concepts and developments. ADFL Bulletin, 18,1: 9-10.

Clark, John L.D., and Ray Clifford. 1987. The FSI/ILR/ACTFL proficiency scales and testing techniques: development, current status, and needed research. In Valdman, 1-18.

Cummins, Patricia. 1987. School-college articulation and proficiency standards: a status report. ADFL Bulletin, 19, 8-15.

Engber, Cheryl. 1987. Summary of the discussion session. In Valdman, 55-8.

Freed, Barbara. 1987. Preliminary impressions of the effects of proficiency-based language requirements. Foreign Language Annals, 20,2, 139-146.

Frith, James R. 1979. Testing the FSI testing kit. ADFL Bulletin, 11,2:12-14.

Galloway, Vicky. 1980. Perceptions of the communicative efforts of American students of Spanish. Modern Language Journal, 64:428-33.

Gannon, R E. 1980. Appropriateness and the foreign language learner. English Language Teaching Journal, 34, 90-3.

Gaudiani, Claire. 1987. Dissemination and diffusion of progress in the foreign language teaching field: a national plan. In Lambert 163-176.

Henning, Grant. 1983. Oral proficiency testing:comparative validities of interview, imitation, and completion models. Language Learning, 33,3:315-32.

Higgs, Theodore. ed. 1984. Teaching for proficiency; the organizing principle. Lincolnwood; National Textbook Company.

Hiple, David, and Joan Manley. 1987. Testing how well foreign language teachers speak: a state mandate. Foreign Language Annals, 20,2, 147-153.

Hummel, Robert D. 1985. Evaluating proficiency in comprehension skills: how can we measure what we can't observe ? ADFL Bulletin, 16,2: 13-16.

Kramsch, Claire. 1986. Proficiency versus achievement: reflections on the proficiency movement. ADFL Bulletin, 18,1: 22-4.

Lambert Richard, ed. 1987. Foreign language instruction: a national agenda. Annals of the American Academy of Political and Social Science. Newbury Park: Sage Publications.

Lantolf, James P., and William Frawley. 1985. Oral Proficiency testing; a critical analysis. Modern Language Journal, 69,4:337-45.

1988. Proficiency: understanding the construct. Studies in Second Language Acquisition 10, 181-95.

Liskin-Gasparro, Judith. 1984a. The ACTFL Proficiency Guidelines: A Historical Perspective. In Higgs, 11-42.

1984b. The ACTFL proficiency guidelines:gateway to testing and curriculum. Foreign Language Annals, 17,5:475-89.

Loveday, Leo. 1982. The sociolinguistics of learning and using a non-native language. Oxford, Pergamon Press.

Ludwig, Jeanette. Native speaker judgments of second language learners' efforts at communication: A review. Modern Language Journal 66: 274-83, 1984.

Magnan, Sally. 1987. Legal caveats on communicative proficiency testing for graduation requirements or teacher certification. In Valdman, 45-50.

Morello, Joseph. 1988. Attitudes of students of French toward required language study. Foreign Language Annals, 21,5: 435-42.

Nichols, Johanna. 1988. Language study, international study, and education. Profession 88 (MLA) 10-17.

Omaggio, Alice. 1986. Teaching language in context. Boston;Heinle & Heinle.

Reschke, Claus. 1985. Proficiency based foreign language teacher certification: the Texas Project. ADFL Bulletin, 16,2: 17-21.

Savignon, Sandra. 1985. The evaluation of communicative competence: the ACTFL provisional proficiency guidelines. Modern Language Journal, 69, 2, 129-34.

Schulz, Renate 1988. Proficiency-based foreign language requirements: a plan for action. ADFL Bulletin, 19, 2: 24-8.

Shohamy, Elana. 1983. Rater reliability of the oral interview speaking test. Foreign Language Annals, 16,3:219-22.

Valdman, Albert ed. 1987. Proceedings of the Symposium on the evaluation of foreign language proficiency. Indiana University, Bloomington.